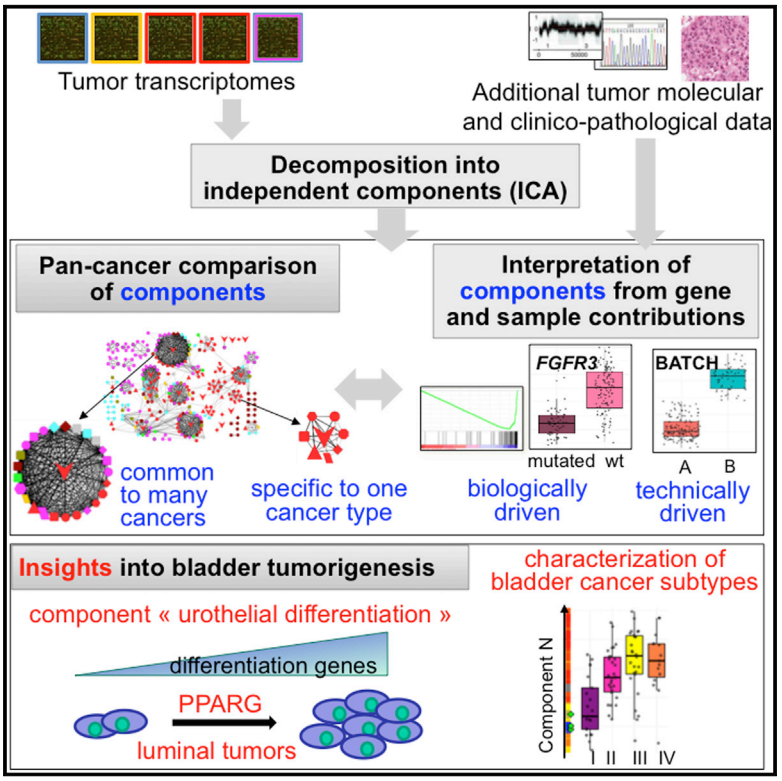
Report

Cell Reports

Independent Component Analysis Uncovers the Landscape of the Bladder Tumor Transcriptome and Reveals Insights into Luminal and Basal Subtypes

Graphical Abstract



Highlights

ICA analysis uncovers the various factors influencing transcriptomic data

ICA of multiple data sets reveals cancer-shared and cancer-type-specific signals

The components identified by ICA allow characterization of bladder tumor subtypes

PPARG is a protumorigenic gene associated with differentiation in bladder cancer

Authors

Anne Biton, Isabelle Bernard-Pierrot, ..., Andrei Zinovyev, François Radvanyi

Correspondence

francois.radvanyi@curie.fr

In Brief

Extracting biological insights from large-scale data is both a challenge and an opportunity. Biton et al. now analyze bladder tumor transcriptomes. An enrichment analysis of contributing genes combined with molecular and clinical annotations of tumor samples identifies biologically relevant components, some of which are shared with other carcinoma types.

Biton et al., 2014, Cell Reports 9, 1235–1245
November 20, 2014 ©2014 The Authors
http://dx.doi.org/10.1016/j.celrep.2014.10.035

Cell Reports
Report

Independent Component Analysis Uncovers the Landscape of the Bladder Tumor Transcriptome and Reveals Insights into Luminal and Basal Subtypes

Anne Biton,[1,2,3,4,16,18] Isabelle Bernard-Pierrot,[1,4,16] Yinjun Lou,[1,4,19] Clémentine Krucker,[1,4] Elodie Chapeaublanc,[1,4] Carlota Rubio-Pérez,[5] Nuria López-Bigas,[5,6] Aurélie Kamoun,[1,4] Yann Neuzillet,[1,4,7,8] Pierre Gestraud,[1,2,3] Luca Grieco,[1,2,3] Sandra Rebouissou,[1,4,20] Aurélien de Reyniès,[9] Simone Benhamou,[10,11] Thierry Lebret,[7,8] Jennifer Southgate,[12] Emmanuel Barillot,[1,2,3] Yves Allory,[13,14,15] Andrei Zinovyev,[1,2,3,17] and François Radvanyi[1,4,17,*]
[1]Institut Curie, Centre de Recherche, 75248 cedex 05 Paris, France
[2]INSERM, U900, 75248 cedex 05 Paris, France
[3]Ecole des Mines ParisTech, 77305 cedex Fontainebleau, France
[4]CNRS, UMR 144, Oncologie Moléculaire, Equipe Labellisée Ligue Contre le Cancer, Institut Curie, 75248 cedex 05 Paris, France
[5]Research Unit on Biomedical Informatics, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, 08003 Barcelona, Spain
[6]Catalan Institution for Research and Advanced Studies (ICREA), 08003 Barcelona, Spain
[7]Service d'Urologie, Hôpital Foch, 92150 Suresnes, France
[8]Université de Versailles - Saint-Quentin-en-Yvelines, Faculté de Médecine Paris - Ile-de-France Ouest, 78280 Guyancourt, France
[9]Ligue Nationale Contre le Cancer, Cartes d'Identité des Tumeurs Program, 75013 Paris, France
[10]CNRS, UMR 8200, Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France
[11]INSERM, U946, 75010 Paris, France
[12]Jack Birch Unit of Molecular Carcinogenesis, Department of Biology, University of York, York Y010 5DD, UK
[13]AP-HP, Département de Pathologie, Hôpitaux Universitaires Henri Mondor, 94000 Créteil, France
[14]INSERM, U955, 94000 Créteil, France
[15]Université Paris-Est, Faculté de Médecine, 94000 Créteil, France
[16]Co-first author
[17]Co-senior author
[18]Present address: Department of Medicine, Lung Biology Center, University of California San Francisco, San Francisco, CA 94143-2922, USA
[19]Present address: ZheJiang University, Institute of Hematology, HangZhou, ZheJiang 310058, China
[20]Present address: INSERM, UMR-674, Génomique Fonctionnelle des Tumeurs Solides, IUH, 75010 Paris, France
*Correspondence: francois.radvanyi@curie.fr
http://dx.doi.org/10.1016/j.celrep.2014.10.035

SUMMARY

Extracting relevant information from large-scale data offers unprecedented opportunities in cancerology. We applied independent component analysis (ICA) to bladder cancer transcriptome data sets and interpreted the components using gene enrichment analysis and tumor-associated molecular, clinicopathological, and processing information. We identified components associated with biological processes of tumor cells or the tumor microenvironment, and other components revealed technical biases. Applying ICA to nine cancer types identified cancer-shared and bladder-cancer-specific components. We characterized the luminal and basal-like subtypes of muscle-invasive bladder cancers according to the components identified. The study of the urothelial differentiation component, specific to the luminal subtypes, showed that a molecular urothelial differentiation program was maintained even in those luminal tumors that had lost morphological differentiation. Study of the genomic alterations associated with this component coupled with functional studies revealed a protumorigenic role for PPARG in luminal tumors. Our results support the inclusion of ICA in the exploitation of multiscale data sets.

INTRODUCTION

Large-scale projects (e.g., Cancer Genome Project [CGP], The Cancer Genome Atlas [TCGA], or the International Cancer Genome Consortium [ICGC]), and the efforts of individual laboratories, are generating massive amounts of high-throughput molecular data often associated with clinicopathological characteristics. Transcriptome data for tumors are the most commonly available type of large-scale molecular data. These remain difficult to interpret because the transcriptome is influenced by various overlapping biological factors linked to the tumor cells or to the tumor microenvironment and nonbiological factors linked to sample processing or data generation. The need to deconvolute these factor effects led to the use of methods originally developed to solve the blind source separation problem (Jutten and Hérault, 1991), with the aim of recovering hidden signal

sources from the observed output mixture. Independent component analysis (ICA) is one of these methods. ICA models the level of expression of each gene in a given sample as a linear weighted sum of several independent components, where each component captures the effect of one of the factors/processes. The expression data matrix is thus decomposed into a number of components, each of which is characterized by an activation pattern both across genes and across samples. The genes with the largest projection onto a component (providing the greatest contribution) are the genes the most strongly influenced by the process associated with this component. The contribution value of a sample reflects the activity of the component in this sample. Most of the studies applying ICA to cancer transcriptome data have focused on the interpretation of components based on the contribution of the genes to the components (Liebermeister, 2002; Lee and Batzoglou, 2003; Carpentier et al., 2004; Chiappetta et al., 2004; Frigyesi et al., 2006; Teschendorff et al., 2007). No study, to our knowledge, has interpreted ICA results taking into account additional data associated with the samples, such as clinical and pathological data (including histopathological images), mutations and copy number alterations, and conditions of procurement and processing of the samples.

Bladder cancer is one of the most common cancers in North America and Europe (Ferlay et al., 2010). The stage classification differentiates between non-muscle-invasive (CIS [carcinoma in situ], Ta, T1) and muscle-invasive tumors (T2, T3, T4). Gene expression clustering is commonly used to identify subtypes in cancer with the aim of gaining new insights into the molecular heterogeneity of tumors and of improving the management of cancer patients. Recently, basal and luminal subtypes of muscle-invasive bladder cancer that shared molecular features with basal and luminal breast cancers have been identified (Sjödahl et al., 2013; Choi et al., 2014; Cancer Genome Atlas Research Network, 2014; Damrauer et al., 2014; Rebouissou et al., 2014).

We show here that by applying ICA to bladder cancer transcriptome data sets, exploiting additional molecular and clinicopathological data, and comparing to ICA applied to multiple cancer types, it is possible to obtain insights into bladder carcinogenesis.

RESULTS

ICA Applied to the CIT Bladder Cancer Transcriptome Data Set: Interpretation of Components Based on Gene Projections

ICA was applied to our reference data set, which was produced in the framework of the French national Cartes d'Identité des Tumeurs (CIT) program from a series of 198 tumors and three normal samples (CIT series). These 198 tumors comprised both non-muscle-invasive tumors (n = 106) (deposited in ArrayExpress, accession number E-MTAB-1940) and muscle-invasive tumors (n = 92) (E-MTAB-1803, Rebouissou et al., 2014). Twenty independent components (ICs) were computed.

The first interpretation of each component was based on analysis of the genes with the largest projections on the component (the contributing genes, the most strongly influenced by the process associated with this component; Table S1). We investigated the association of the component with a specific cell type, known pathways, or sets of coregulated genes, by studying the enrichment of these contributing genes in predefined gene sets (such as Gene Ontology, KEGG, or BioCarta, see Supplemental Experimental Procedures). We also investigated whether these genes were located in a particular chromosomal region or could be distinguished in terms of sequence features, such as GC content.

Eleven components were associated with a biological process on the basis of enrichment analysis (Tables 1, S2 and S3, sheet "Association_ICs_genomicLocation"). Among these components, we were able to distinguish between components associated with the tumor cells and components associated with the tumor microenvironment. Three of the tumor cell-driven components were linked to various types of differentiation (squamous and basal differentiation [CIT-6], urothelial differentiation [CIT-9], neural differentiation [CIT-18]), one was associated with the cell cycle (CIT-7), one was associated with mitochondria (CIT-4), and another (CIT-10) contained six neighboring genes located at 11p15.5, including the *IGF2/H19* imprinted locus (Figure S1A). Three components were associated with the stroma: CIT-3 was assigned to smooth muscle, CIT-8 to the immune response mediated by B and T lymphocytes, and CIT-12 to tumor-associated myofibroblasts. Two biologically interpretable components could not be attributed to either tumor cells or the stroma (CIT-5, CIT-14). CIT-5 was enriched in interferon response genes and CIT-14 was enriched in inflammatory and early response genes. One component, CIT-2, was not significantly enriched in any particular gene set, but gene projections onto this component were strongly correlated with GC content of the corresponding genes (Figure S1B).

ICA Applied to the CIT Bladder Cancer Transcriptome Data Set: Interpretation of the Components Based on Contributions of the Tumor Samples

We further interpreted the components by studying the association of each component with specific tumor sample features by analyzing the values of sample contributions to the components (Table S3, sheet "Sample_Contributions", and Figure S1).

The samples of the CIT series were comprehensively annotated (Table S3, sheet "Sample_Annotations") for clinical and pathological data, the mutational status of the genes most commonly mutated in bladder cancer, genomic alterations assessed by CGH array, transcriptomic signatures (the carcinoma in situ signature (Dyrskjøt et al., 2004), and the multiple regional epigenetic silencing phenotype (MRES phenotype) (Vallot et al., 2011)) and technical factors.

We investigated the distributions of the tumors of the CIT series on the components as a function of these parameters (Table S3, sheet "Association_Annotation_ICs"). This approach enabled us to interpret two additional components (CIT-1 and CIT-13) and deepened our understanding of two components previously assigned based on gene contributions (CIT-14 and CIT-18).

CIT-13 was associated with the two well-known pathways of bladder tumor progression (Figure 1A). Indeed, most of the tumors displaying a negative contribution to CIT-13 were Ta

Table 1. Interpretation of the Independent Components of the CIT Bladder Cancer Data Set Using the Contributing Genes

Source	Component ID	Component Name	Contributing Genes (n)	Interpretation of the Contributing Genes
Tumor cells	CIT-6	basal-like	151	enriched in gene sets "keratinization," "keratinocyte differentiation"; markers of epithelial basal cells: KRT5, KRT6A, KRT14, KRT16, KRT17, DSC3, etc.
	CIT-4	mitochondria	61	enriched in mitochondrial genes
	CIT-7	cell cycle	187	enriched in targets of E2F and genes involved in cell cycle
				markers of proliferation: TOP2A, CDC20, CCNB2, CDK1, BUB1B, AURKA, kinesins (KIF4A, KIF2C, KIF14, KIF15, KIF11, KIF20A, KIF23, KIFC1), etc.
	CIT-9	urothelial differentiation	181	markers of urothelial differentiation: UPK1A, UPK1B, UPK2, KRT20, PPARG, BAMBI. TFs involved in differentiation: FOXA1, GATA3, GRHL3
	CIT-10	11p15.5	163	five genes colocalized at 11p15.5 (*H19*, *SYT8*, *TNNT3*, *TNNI2*, *LSP1*)
	CIT-13	bladder cancer pathways[a]	175	
	CIT-18	neuroendocrine[b]	122	enriched in neuronal genes: *CELSR2*, *CELSR3*, *ENO2*, *EMX2*, *NMU*, etc.
Stroma	CIT-3	smooth muscle	197	enriched in targets of the serum response factor SRF, and in gene sets associated with smooth muscle
				markers of the smooth muscle: calponin 1 (CNN1), synemin (SYNM), actins (ACTG2, ACTC1, ACTA2), myosins (MYL9, MYLH11), desmin (DES), transgelin (TAGLN), etc.
	CIT-8	lymphocytes B&T (LB&T)	211	enriched in gene sets associated with immune reaction
				markers of lymphocytes (immunoglobulins, chemokines, members of the MHC class II, CD79A, POU2AF1, and markers of T cells CD2, CD53, CD3D, CD8A, GZMA, GZMB, LCP2, TNFRSF17, etc.)
	CIT-12	myofibroblasts	192	enriched in gene sets associated with extracellular matrix
				markers of myofibroblasts: periostin (POSTN,) lumican (LUM), decorin (DCN), collagens (COL1A1, COL1A2, COL5A1, COL5A2, COL6A2, COL6A3, COL15A1, COL3A1, COL12A1), etc.
Tumor cells or stroma	CIT-5	interferon response (IFN)	169	enriched in targets of interferon regulatory factors IRF1, IRF2, IRF7, IRF8, and the TF STAT1. Among the first 25 contributing genes, 11 encode interferon-induced proteins *IFIT1*, *IFIT3 IFI6*, *IFI27*, *IFI35*, *IFI44*, *IFI44L*, *MX1*, *MX2*, *IFI6*, *IFIH1*. *IRF9* is also a contributing gene. The presence of *MX1* and *CXCL10* in the contributing genes indicates a type I interferon response.
Involvement of technical factors	CIT-1	batch effect[a]	147	
	CIT-2	GC content	32	the projection values of all genes are highly correlated to their GC content (Pearson correlation = 0.66)
	CIT-14	inflammation/early response/type of surgery[b]	173	enriched in genes involved in inflammatory response and response to stress
				contributing genes including chemokine genes (*CXCL1*, *CXCL2*, *CXCL3*), early response to stress genes (*EGR1*, *EGR2*, *EGR3*, *IER3*), interleukin genes (*IL8*, *IL1A*), *FOS*, *FOSB*, *DUSP1*, and *DUSP5*

See also Figure S1 and Tables S1 and S2.
[a]Assignation of components using tumor contribution.
[b]Further assignation of components using tumor contribution.

papillary tumors, frequently mutated for *FGFR3* (Figure 1B), the marker of the Ta pathway. Conversely, the tumors displaying a positive contribution to CIT-13 were mostly T2-4 tumors, with high rates of *TP53* mutation, expressing markers preferentially associated with the CIS pathway (the CIS signature and the MRES phenotype) (Figure 1B). T1 tumors were found on both sides of the component (Figure 1B). The distribution of the tumors on CIT-13 was therefore entirely consistent with the definition of the two pathways of bladder tumor progression. CIT-1 was not associated with any biological information. The traceability of processing for the CIT samples made it possible to attribute this component to a batch effect (Figure S1C).

The strongest association for component CIT-14 was with the type of surgery (transurethral resection of the bladder tumor versus cystectomy) ($p = 1.89 \times 10^{-17}$, Table S3, sheet "Association_Annotation_ICs"; Figure S1D). The enrichment of the contributing genes of CIT-14 in early response genes (Table 1) may therefore in part reflect the different stresses induced by the two different surgical procedures. The enrichment of CIT-14 in genes previously reported as differentially expressed in

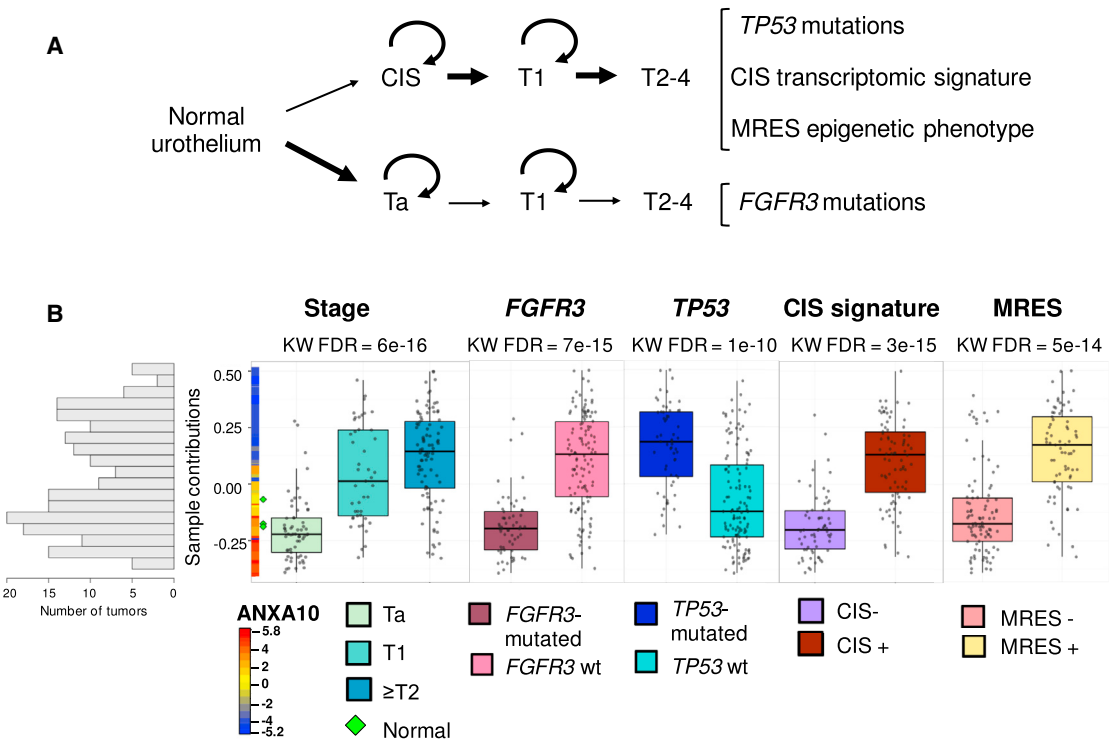


Figure 1. Association of the CIT-13 Component with the Two Pathways of Bladder Tumor Progression

(A) The two pathways of bladder tumor progression (Ta pathway and CIS pathway). Ta tumors constitute the largest group of bladder tumors (50%) at first diagnosis. Ta tumors often recur but only rarely progress via T1 tumors to muscle-invasive tumors (T2–T4 tumors). The Ta pathway is characterized by a high frequency of *FGFR3* mutations. The other pathway corresponds to the carcinoma in situ (CIS) pathway. CIS often progress (in about 50% of cases) to T1 and then to muscle-invasive tumors. This pathway is characterized by early *TP53* mutations (Spruck et al., 1994), a specific transcriptomic signature (Dyrskjøt et al., 2004), and an epigenetic phenotype (MRES, Vallot et al., 2011). For box plots, the bottom, top, and middle bands of the boxes indicate the 25th, 75th, and 50th percentiles, respectively. Whiskers extend to the most extreme data points no more than 1.5 interquartile range from the box.

(B) Distribution of the tumors on the CIT-13 component as a function of pathological and molecular characteristics. The histogram of the sample contributions to CIT-13 is shown on the left. The sample contribution values are given on the vertical axis. On this axis is also shown the expression level of one of the strongest contributing genes (*ANXA10*) of the IC, each point representing one tumor sample, and its color indicating the relative level of expression of the gene. The color scale for the expression level of *ANXA10* is shown below. The box plots and superimposed individual data points show the distribution of the samples on CIT-13 as a function of stage (Ta, T1, and tumors of higher stage: T2–T4), *FGFR3* and *TP53* mutation status, the CIS transcriptomic signature (Dyrskjøt et al., 2004), and the multiple regional epigenetic silencing (MRES) phenotype (Vallot et al., 2011). The normal samples are represented as green diamonds. False discovery rates (FDRs) for Kruskal-Wallis tests are provided at the top of each plot.

See also Table S3.

prostate cancers as a function of sample type (biopsy versus prostatectomy specimens) (Lin et al., 2006) supports this conclusion (Table S2, sheet 1). We cannot exclude the possibility that a biological process not directly related to the type of surgery may also contribute to this component, which we refer to as the inflammation/early response/type of surgery. The CIT-18 component, which was enriched in neural genes (Table 1), was due to the presence of neuroendocrine tumors in the CIT series (Figure S1E).

Representative pathological sections were available for the CIT tumor series. For each component, the sections were ranked according to the contribution of the corresponding tumor sample to the component (http://microarrays.curie.fr/publications/UMR144/LuminalBasalBladder/) and were reanalyzed in this context. This confirmed the assignment of components CIT-3, CIT-12, and CIT-7 to smooth muscle, myofibroblasts, and cell cycle, respectively (Figure S2).

ICA of Multiple Transcriptome Data Sets, Bladder, and Other Types of Cancer Allows Distinguishing Reproducible Cancer Type-Specific Components and Components Common to Different Cancer Types

We investigated the reproducibility of the components retrieved from our reference bladder tumor set (CIT series), by applying ICA to seven other bladder cancer transcriptome data sets, six series of breast carcinomas and eight series of carcinomas of various types (colon, rectum, endometrium, kidney, lung squamous cell and lung adenocarcinoma, ovary, prostate). The data sets, their accession numbers, and gene projections for all data sets are given in Table S4.

We used a correlation graph approach to compare the ICs across the different data sets (see Supplemental Experimental Procedures) (Figure 2). In the correlation graph, each node represents a component computed for a data set, and each edge connects two components if the corresponding gene projections

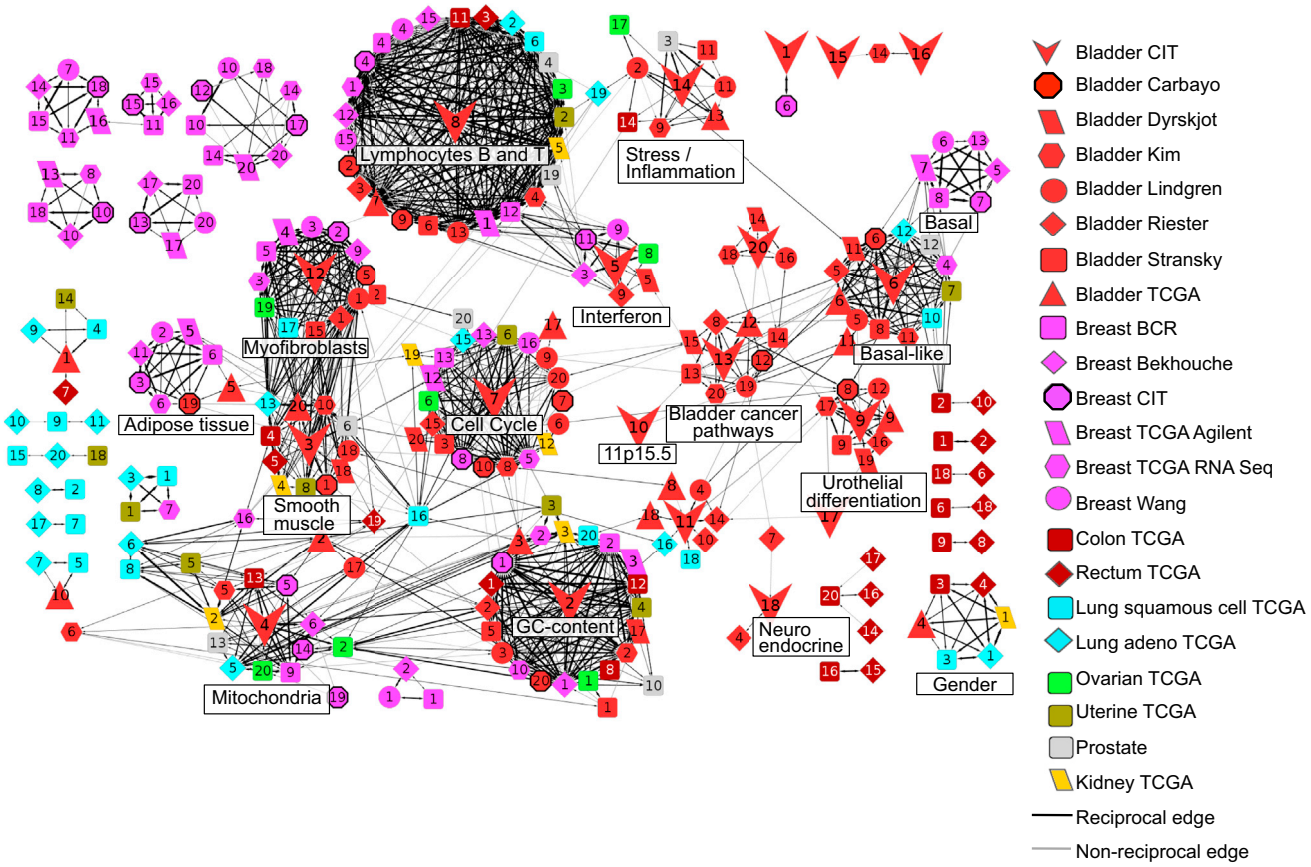


Figure 2. Comparison of the Independent Components across Data Sets of Various Cancer Types Using Correlation-Based Graphs

Correlation-based graph of the independent components (ICs) calculated for seven bladder cancer data sets plus 14 other gene expression data sets of different types of cancer, including six breast carcinoma data sets and one data set for each of the following cancer types: prostate, colon, and rectum adenocarcinomas, uterine corpus endometrioid carcinoma, kidney (clear cell carcinomas), lung (squamous cell carcinomas and adenocarcinomas), and ovarian serous adenocarcinomas. Each node denotes an IC; the node colors indicate the type of cancer, whereas their shapes distinguish between data sets of the same cancer type. The numerical IDs indicate the order in which the ICs were returned by the icasso method. The graphs were visualized with Cytoscape (Cline et al., 2007). The edge thickness indicates the degree of correlation between the two ICs linked by the edge. Only pairs of ICs with absolute correlation values exceeding 0.35 are shown. Reciprocal edges are colored black, whereas non-reciprocal edges are in gray. Reproducible components appear on the graph as a cluster of tightly interconnected nodes. See also Tables S4 and S5.

are correlated. Clusters of components (pseudocliques, see Supplemental Experimental Procedures) were observed in the structure of the correlation graph (the genes contributing to the components of each clique are given in Table S5). Some clusters were composed of densely interconnected nodes and included components common to most types of cancer (immune response mediated by B and T lymphocytes, cell cycle, GC content, myofibroblasts/smooth muscle). Other clusters were common to some, but not all types of cancer (including gender, basal-like, mitochondria, interferon response, inflammation/early response genes/type of surgery-associated clusters). Because we used several data sets for bladder and breast cancers, we were able to identify pseudocliques specific to these cancers. Two bladder-cancer-specific pseudocliques consisted of components found in all bladder cancer data sets: one was associated with urothelial differentiation, whereas the other was associated with the two pathways of bladder tumor progression.

Association of the Recently Identified Expression-Based Subtypes of Muscle-Invasive Bladder Cancer with Biological Processes

Muscle-invasive bladder cancers are clinically and biologically heterogeneous. Recently, expression-based clusters that resembled basal and luminal subtypes of breast cancer have been reported in muscle-invasive bladder cancer (Sjödahl et al., 2013; Cancer Genome Atlas Research Network, 2014; Choi et al., 2014; Damrauer et al., 2014; Rebouissou et al., 2014). To characterize these muscle-invasive bladder cancer subtypes, we compared their distribution on the different components, which had been interpreted in the CIT series. We studied the four subtypes from the TCGA classification (Cancer Genome Atlas Research Network, 2014): cluster I (papillary luminal), cluster II (luminal), cluster III (basal-like), and cluster IV. These four clusters were identified in the CIT series using a centroid-based predictor (see Supplemental Experimental Procedures). The six components displaying the most significant

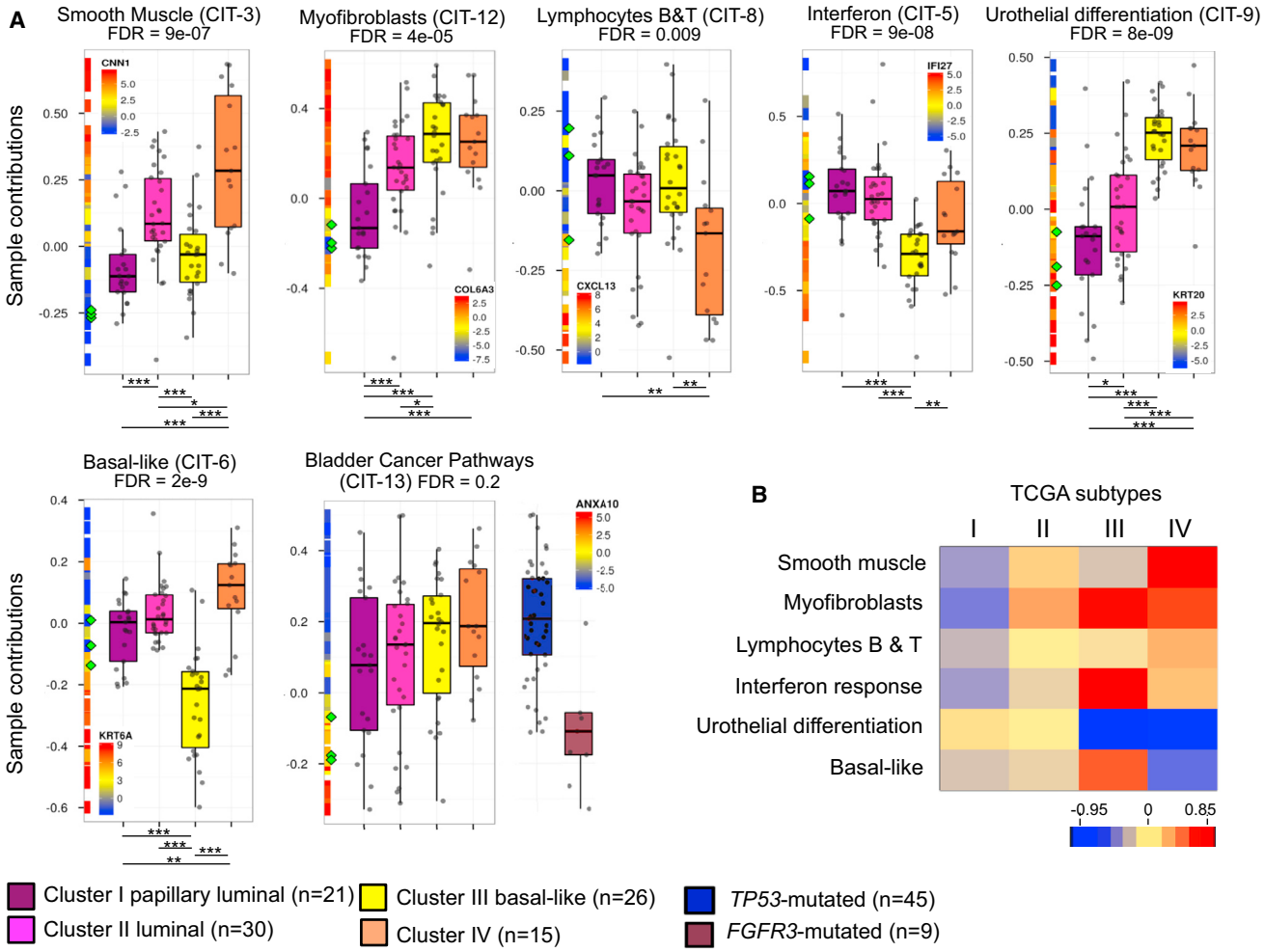


Figure 3. Characterization in the CIT Series of the Four Gene Expression Subtypes as Defined in the Bladder Cancer TCGA 2014 Classification

Cluster I (papillary luminal, in dark magenta), cluster II (luminal, in magenta), cluster III (basal-like, in yellow), and cluster IV (orange). The six components displaying the most significant differences between these subtypes as well as the progression pathway component are shown.

(A) Each box plot represents the distribution of a subgroup of tumors on the IC and is overlaid with the individual data points. The vertical lines at the left of the box plots indicate the expression level of one of the strongest contributing genes of the IC, each point representing one tumor sample, and its color indicating the relative level of expression of the gene. For box plots, the bottom, top, and middle bands of the boxes indicate the 25th, 75th, and 50th percentiles, respectively. Whiskers extend to the most extreme data points no more than 1.5 interquartile range from the box. The diamond-shaped green points indicate the positions of the normal samples. FDR values for Kruskal-Wallis tests are provided at the top of each plot. The p values (Wilcoxon test) for each pairwise comparison are shown at the bottom of each plot: *$0.01 \leq p < 0.05$; **$0.001 \leq p < 0.01$; ***$p < 0.001$.

(B) Heatmap representing the median of the sample contributions of each subtype to the components. For the purpose of this figure, the vector of sample contributions of some components (lymphocytes, interferon response, urothelial differentiation, basal-like) has been reversed such that the over- (respectively, under-) expression of the most contributing genes of the components are on the positive (respectively, negative) side of the component and thus appear in red (respectively, blue) color.

See also Figure S2 and Table S3.

differences between these clusters and the progression pathway component are shown in Figure 3A. As expected, basal-like tumors presented a basal differentiation (their distribution on the basal-like component [CIT-6] corresponded to high expression of basal markers, such as KRT6A), and a low level of urothelial differentiation (their distribution on the urothelial differentiation component (CIT-9) corresponded to low expression of urothelial differentiation markers, such as KRT20). We also showed that this subtype presented a strong interferon response (as shown by its distribution on the interferon-response component [CIT-5]). The distribution of basal-like tumors on the component associated with the bladder cancer pathways (CIT-13) was similar to that of *TP53*-mutated tumors and different from that of *FGFR3*-mutated tumors, suggesting that basal-like tumors belonged to the CIS pathway (see diagram in Figure 1A). Their distribution on the components associated with stroma (CIT-3 [smooth muscle component], CIT-8 [lymphocytes T and B component], and CIT-12 [myofibroblast component]) indicated

that basal-like tumors were relatively poor in smooth muscle cells and lymphocytes T and B and relatively rich in myofibroblasts (strong expression of myofibroblast markers). The two luminal subtypes (clusters I and II) presented a high level of differentiation. Indeed, their distribution on the urothelial differentiation component (CIT-9) corresponded to high expression of urothelial differentiation markers, such as KRT20. These two subtypes differed in terms of stroma, cluster II was richer in smooth muscle (CIT-3) and myofibroblasts (CIT-12). A schematic representation summarizing the characteristics of each subtype according to the components is shown in Figure 3B.

Genetic and Functional Evidence for a Role of PPARG in Urothelial Differentiation and in Bladder Carcinogenesis

The urothelial differentiation component (CIT-9), found in all bladder cancer data sets studied, was specifically associated with bladder luminal tumors. We studied this component in more detail.

Most luminal tumors (subtypes I and II) were distributed on the differentiated side of the component (high expression of differentiation markers such as *KRT20*, *GRHL3*, *UPK1A*, *BAMBI*, and *PPARG*) (Figure 4A). As expected, Ta low-grade tumors (G1/G2) were on the differentiated side of the component (Figure 4A) and were differentiated at the morphological level (Figure 4B; http://microarrays.curie.fr/publications/UMR144/LuminalBasalBladder/). Surprisingly the tumors with the highest levels of urothelial differentiation markers (most negative contributions to the component) were high-grade tumors (G3) (Figure 4A). By morphology, these tumors both lacked architectural organization and displayed unpolarized and atypical nuclei (Figure 4B; http://microarrays.curie.fr/publications/UMR144/LuminalBasalBladder/). These observations indicate that there was no relationship between molecular differentiation measured on this component and morphological grade assessed by architectural organization and nuclear atypia.

To identify protumorigenic genes potentially involved in the carcinogenesis of luminal tumors, we looked for altered genomic regions associated with the urothelial differentiation component. Among these regions, we selected the regions of gains that were found in tumors associated with the differentiated side of the component and that contained a contributing gene associated with differentiation. In the CIT series, a region of genomic alteration (chr3:1650731-12883379) encompassing *PPARG*, a contributing gene associated with urothelial differentiation (Table S6) was the most significantly associated with the component. The alterations observed in this region were mostly gains (in 51 of the 178 tumors, 29%; with losses observed in seven of the 178 tumors, 4%) (Figure 4C). Computation of the minimal region of gain (Rouveirol et al., 2006) (see Supplemental Experimental Procedures) identified a 774 kb region containing nine genes, including *PPARG*. These genomic alterations of *PPARG* were correlated with its expression in both tumors and bladder tumor-derived cell lines (Figure S3A). The use of high-resolution SNP arrays (Illumina Human 610) for 11 tumors presenting a gain of this region made it possible to narrow down the minimal region of gain to 273.9 kb including only *PPARG*. The distribution of the tumors of the TCGA series on the urothelial differentiation component according to genomic alterations confirmed that *PPARG* gains were associated with differentiated tumors (Figure S3B). These findings provide genetic evidence that PPARG was involved in tumorigenesis of these tumors. Consistent with this, pathway enrichment analysis of the contributing genes of the urothelial differentiation component showed enrichment in genes of the PPAR pathway: *PPARG*, *FABP4*, *HMGCS2*, *ACSL5*, *ACOX1*, *APOA2*, and *ACOX3* (Table S2, sheet "CIT9-GOstat"). To demonstrate the functional involvement of PPARG in bladder tumors, we studied the effect of small interfering RNA (siRNA)-mediated *PPARG* knockdown on the growth of nine bladder-cancer-derived cell lines with different levels of *PPARG* mRNA. Two different siRNAs targeting *PPARG* mRNA were used, and a downregulation of 65%–90% of *PPARG* mRNA was obtained (Figure S3C). The largest decrease in cell viability (50%) was observed for the two cell lines presenting a genomic alteration of *PPARG* (amplification for UMUC9 and gain for SD48) (Figure 4D, left panel). Overall, *PPARG* mRNA level was significantly correlated with the effect of siRNA-mediated *PPARG* knockdown, because higher levels of *PPARG* expression were associated with larger decreases in cell viability (Figure 4D, right panel). We also investigated the effect of *PPARG* knockdown on the cloning efficiency of three cell lines (SD48, UMUC9, and MGHU3). The inhibition of growth in the soft agar assay was similar (60% for cell lines with *PPARG* genomic alteration, SD48, and UMUC9, and 20% for MGHU3) to the decrease in cell viability in 2D culture (Figure 4E). These results indicated that PPARG was involved in tumor cell growth in PPARG expressing bladder cancer cells.

PPARG is a transcription factor of the nuclear receptor type 1 subfamily and has been implicated in urothelial differentiation (Varley et al., 2004). The presence of PPAR target genes among the contributing genes of the urothelial differentiation component suggested that PPARG could control the expression of several contributing genes of this component. We tested this hypothesis by comparing the transcriptome of the SD48 cell line treated with three different siRNAs targeting *PPARG* or with the transfection reagent (Lipofectamine). We found that 198 genes were significantly deregulated by siRNA treatment (107 genes significantly downregulated and 91 genes upregulated) (Table S7). This set of deregulated genes was significantly enriched (Fisher's test, $p < 0.0001$) in genes contributing to the CIT urothelial differentiation component (21 of 198; Figure 4F; Table S7). Most of the 21 genes of the urothelial differentiation component controlled by *PPARG* were upregulated by this transcription factor (downregulation observed upon *PPARG* knockdown for 18 of these 21 genes) and, as expected, were associated with differentiation (17 of these 18 genes). Conversely, two of the three genes downregulated by *PPARG* (upregulated following *PPARG* knockdown) were associated with undifferentiated tumors.

DISCUSSION

We applied independent component analysis (ICA) to 21 transcriptome data sets, including eight data sets for bladder cancer and five for breast cancer. For one of the bladder cancer data sets (the CIT data set that was generated by our group), clinical, pathological, molecular, and sample processing data were available. The use of these additional data made it possible to assign

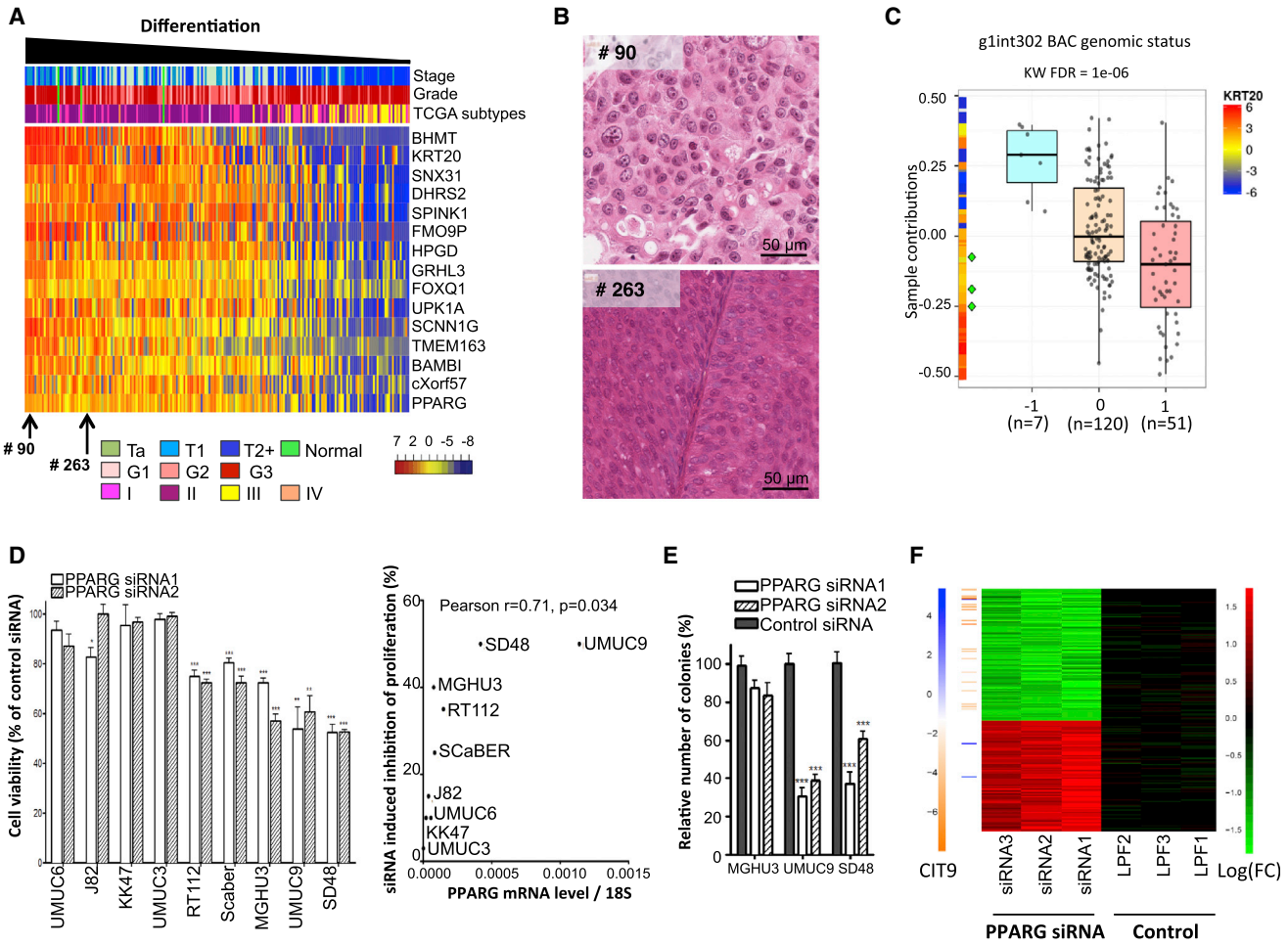


Figure 4. Identification of *PPARG* as an Oncogene Controlling Cell Viability and Differentiation in Bladder Tumors

(A) Heatmap of the top 15 contributing genes of CIT-9; the genes (rows) and samples (columns) are ranked according to their contribution to the component. The gene expression profiles are centered. These top 15 most contributing genes include several urothelial differentiation markers (*KRT20*, *GRHL3*, *UPK1A*, *BAMBI*, *PPARG*). The stage, grade, and TCGA tumor subtype classification (I, papillary luminal; II, luminal; III, basal-like; IV) are indicated.

(B) Histopathological sections of two tumors on the differentiated side of the component (high expression of urothelial differentiation markers). Case #90, a T1G3 tumor, despite having one of the highest levels of urothelial differentiation markers lacked any architectural organization and displayed unpolarized and atypical nuclei. Case #263, a TaG2 tumor, which was also on the differentiated side of the component, was differentiated at the morphological level.

(C) Distribution of the tumor samples according to their alteration status for the region on chromosome 3:1650731-12883379 encompassing *PPARG*, as measured by aCGH. FDR value for the Kruskal-Wallis test is provided at the top of the box plot. For box plots, the bottom, top, and middle bands of the boxes indicate the 25th, 75th, and 50th percentiles, respectively. Whiskers extend to the most extreme data points no more than 1.5 interquartile range from the box.

(D) Left: effect of PPARG knockdown on cell viability, as assessed by MTT assays, of nine bladder-cancer-derived cell lines. Two different siRNAs were used. Right: correlation between cell viability inhibition induced by *PPARG* knockdown (mean of the two different siRNAs) and *PPARG* expression.

(E) Effect of *PPARG* knockdown on anchorage-independent colony formation in soft agar. The values reported are the means ± SD of three independent experiments carried out in triplicate.

(F) Heatmap visualization of the 198 regulated transcripts in SD48 cells 96 hr after *PPARG* knockdown by three different siRNAs. The expression ratio (FC) was determined for each gene by comparison of the mean level of expression on lipofectamine control (LPF) chips with that on *PPARG* siRNA chips. The gene expression profiles are centered on the mean level of expression on control chips. For regulated genes contributing to the CIT-9 component with an absolute value for the contribution score exceeding 3 (n = 21), the contribution score is indicated according to the color scale on the left side of the heatmap.

See also Figure S3 and Tables S3, S6, and S7.

several components to a biological process or to technical factors and allowed us to characterize subtypes of tumors and to identify molecular mechanisms associated with a component.

Our analysis across different types of cancer identified both components common to some or all cancer types considered, and also components specific to a particular type of cancer. Three components common to all cancer types (B and T lymphocytes, cell cycle, myofibroblasts) probably correspond to the three multicancer gene signatures/attractors (lymphocyte-specific, mitotic chromosomal instability, and mesenchymal transition) identified very recently in breast, colon, and ovarian cancer by a different computational approach (Cheng et al., 2013).

Indeed, many of the genes associated with these components were also found to be markers for these signatures. The smooth muscle and myofibroblasts components were merged in some data sets (the TCGA and Kim et al., 2010 bladder data sets and in the lung adenocarcinoma data set), indicating that the expression of these two stroma components are closely related and are often present in the same samples for these data sets.

The components common to some but not all cancer types included the basal-like component. This component was found in bladder cancer, lung squamous carcinoma, lung adenocarcinoma, and endometrioid carcinoma and was related to a component specific to breast cancer, the basal component. The existence of a basal subgroup in breast cancer is widely recognized (Bertucci et al., 2012; Sørlie et al., 2001). It is becoming clear that basal-like subgroups also exist in other cancer types (Chung et al., 2004; Wilkerson et al., 2010; Di Palma et al., 2012), including, as shown very recently, bladder cancer (Sjödahl et al., 2013; Cancer Genome Atlas Research Network, 2014; Choi et al., 2014; Damrauer et al., 2014; Rebouissou et al., 2014).

TCGA results have shown that most colon and rectum adenocarcinomas cannot be distinguished at the genomic level (Cancer Genome Atlas Network, 2012). Consistent with this observation, adenocarcinomas of the colon and the rectum shared most of their components in our analysis.

The biologically meaningful ICs we identified included four related to tumor microenvironment. A detailed interpretation of these components should help to characterize the tumor microenvironment and the reciprocal interactions between cancer cells and stromal cells.

Luminal tumors differed from the other bladder tumors by their activities on different components: low interferon response and the presence of molecular urothelial differentiation. Surprisingly, molecular urothelial differentiation was maintained even in those luminal tumors that had lost morphological differentiation, as assessed by cytological atypia and the absence of architectural organization. This suggests a dependency on part of the differentiation program for tumor progression. We demonstrated that the transcription factor PPARG positively controls the expression of genes linked to molecular differentiation in tumor cells. In agreement with this, Choi et al. (2014) have recently shown a PPARG activation signature in luminal tumors. In addition, we provide here genetic and functional evidence that this gene is involved in tumorigenesis of molecularly differentiated bladder tumors. Interestingly, the use of pioglitazone, an antitype 2 diabetic drug that is an agonist of PPARG, has been associated with an increased risk of bladder cancer (Turner et al., 2014). This could be due to the protumorigenic role of PPARG, as shown here. A parallel can be drawn between PPARG and another nuclear receptor also associated with differentiation, the estrogen receptor. Indeed, the estrogen receptor is associated with differentiation in the breast, which is necessary for the growth of estrogen receptor-positive tumors, and the use of estrogen increases the risk of breast cancer (Collaborative Group on Hormonal Factors in Breast Cancer, 1997).

There are several advantages of searching for associations between expression and copy number through ICA, as we did here to find *PPARG*, rather than using a classical search for correlation between expression and copy number. The association between a genomic alteration and a component can be directly interpreted in the context of a biological process (represented by the component) and of multiple genes associated with this biological process (the most contributing genes on the component), rather than simply associating one genomic alteration with one gene. Moreover, it reduces the high number of candidates that can be obtained when analyzing the genomic alterations alone. We also found that genomic gains of *GATA3* were associated with the urothelial differentiation component (data not shown). GATA3 is important for the proliferation of luminal breast tumors (Kong et al., 2011) and is therefore likely to play the same role in bladder tumors presenting a molecular urothelial differentiation.

We have developed an R package (*MineICA*, available in Bioconductor: http://www.bioconductor.org/packages/release/bioc/html/MineICA.html) that implements most of the methods used in our study. We also provide a bladder tumor data set that contains not only molecular and anatomoclinical data but also histological data. Integrating image data into multiparametric analyses of tumors is an emerging domain of research. The resource we provide should be useful in this respect.

Our analysis shows that applying ICA to transcriptome data and exploiting clinical, pathological, and genetic alteration data obtained from ongoing large-scale projects will provide insights into cancer biology.

EXPERIMENTAL PROCEDURES

CIT Series: Tissues Samples and Data

Tissues samples of human bladder carcinomas were collected from patients at Henri Mondor Hospital (Créteil, France), Institut Gustave Roussy (Villejuif, France), and Foch Hospital (Suresnes, France). Details of samples and large-scale data are provided in Supplemental Experimental Procedures. All patients provided written informed consent, and the study was approved by the ethics committees of the different hospitals involved in the study (Comité de Protection des Personnes de l'hôpital Henri Mondor, Comité de Protection des Personnes de Boulogne - Ambroise Paré and Comité de Protection des Personnes de Bicêtre). All analyses were performed on the basis of anonymized patient data.

Bioinformatics Analysis

To compute the independent components, we applied the FastICA algorithm (Hyvärinen, 1999) to each data set using *icasso* (Himberg et al., 2004) for selecting stable components. We computed 20 independent components (ICs), using *pow3* nonlinearity and *deflationary* algorithm. We interpreted the genes associated with an IC ("most contributing genes") by performing enrichment analysis with GOstats (Falcon and Gentleman, 2007) and "GSEA Preranked" module (Subramanian et al., 2005). We compared the contributions of predefined groups of tumors to the ICs in Wilcoxon rank-sum and Kruskal-Wallis tests. The ICs were compared across data sets using correlation-based graphs.

Details of computational analyses are provided in Supplemental Experimental Procedures.

Experiments

To study the role of PPARG in bladder-cancer-derived cell lines, we evaluated the effect of PPARG knockdown using siRNAs on cell viability of nine cell lines (J82, KK47, MGHU3, SCaBER, SD48, RT112, UMUC3, UMUC6, UMUC9) and on colony formation in soft agar for three of these cell lines (MGHU3, SD48, UMUC9). PPARG mRNA expression level was determined using quantitative RT-PCR. Genes regulated after PPARG knockdown were determined using Affymetrix DNA microarray. Details of experimental procedures and statistical analysis of the results are provided in Supplemental Experimental Procedures.

ACCESSION NUMBERS

The ArrayExpress accession number for the CIT non-muscle-invasive tumor array data reported in this paper is E-MTAB-1940.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.10.035.

AUTHOR CONTRIBUTIONS

A.B., A.Z., and E.B. designed the computational methodology. A.B. assembled the data and performed the ICA analyses. A.B. implemented tools for data visualization and interpretation in cooperation with F.R. P.G. and E.B. advised on processing the data and tools implementation. A.B., F.R., and A.Z. were responsible for the interpretation of the independent components, helped by I.B.-P., S.R., and J.S. L.G. participated in TCGA data analysis and in the analysis of correlation graph. A.B., A.K., A.Z., and A.d.R. compared CIT and TCGA data. I.B.-P. designed and was responsible for the functional validation studies. C.K., Y.L., Y.N., and I.B.-P. performed the experimental validation studies, and A.K. was involved in their analysis. Y.A., E.C., and I.B.-P. prepared and analyzed histopathological data. Y.A. was responsible for the anatomopathological data and image interpretation, T.L. and S.B. for the clinical data, and C.R.-P., N.L.-P., Y.N., and I.B.-P. for discussing the therapeutic applications of the work. A.B., I.B.-P., F.R., and A.Z. wrote the manuscript. A.Z. supervised the computational work. F.R. coordinated the work. All authors discussed the results and implications and reviewed the manuscript.

ACKNOWLEDGMENTS

This work is part of the "Cartes d'Identité des Tumeurs" (CIT) program funded and developed by the "Ligue Nationale contre le Cancer" (LNCC) (http://cit.ligue-cancer.net). We thank E. Voirin, N. Servant, G. Lucotte, and P. Hupé for their help with bioinformatics data management and analysis. We thank members of the bladder cancer CIT consortium (P. Maillé and D. Vordos, Henri Mondor Hospital; M. Sibony, Cochin Hospital; A. Laplanche, IGR, INSERM; Y. Denoux and V. Molinié, Foch Hospital; E. Letouzé, LNCC) for their constant support. This work was supported by the LNCC (to "Oncologie Moléculaire" and "Computational Systems Biology of Cancer" accredited teams), the Institut Curie (to F.R., E.B., A.Z.), the "Centre National de la Recherche Scientifique" (CNRS) (to F.R.), the "Institut National de la Santé et de la Recherche Médicale" (INSERM) (to E.B., A.Z., S.B., and Y.A.), the INCa (INCa_2960 and 4382 to F.R. and Y.A.), ITMO cancer, systems biology program (to A.Z., E.B., and F.R.), the Labex (no. ANR-10-LBX-0038) part of the IDEX PSL (no. ANR-10-IDEX-0001-02 PSL) (to F.R.), and York Against Cancer (to J.S.). A.B. was supported by a grant from the INCa, from the LNCC, and by NIH grant 5U24 CA143799-04 as part of TCGA project, Y.L. by a grant from the "Fondation Franco-*Chinoise* pour la Science et ses Applications" (FFCSA), Y.N. by a grant from the "Fondation ARC pour la recherche sur le cancer," and A.K. by a grant from the LNCC. The results published here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

Received: January 25, 2014
Revised: August 18, 2014
Accepted: October 13, 2014
Published: November 13, 2014

REFERENCES

Bertucci, F., Finetti, P., and Birnbaum, D. (2012). Basal breast cancer: a complex and deadly molecular subtype. Curr. Mol. Med. *12*, 96–110.

Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330–337.

Cancer Genome Atlas Research Network (2014). Comprehensive molecular characterization of urothelial bladder carcinoma. Nature *507*, 315–322.

Carpentier, A.-S., Riva, A., Tisseur, P., Didier, G., and Hénaut, A. (2004). The operons, a criterion to compare the reliability of transcriptome analysis tools: ICA is more reliable than ANOVA, PLS and PCA. Comput. Biol. Chem. *28*, 3–10.

Cheng, W.-Y., Ou Yang, T.-H., and Anastassiou, D. (2013). Biomolecular events in cancer revealed by attractor metagenes. PLoS Comput. Biol. *9*, e1002920.

Chiappetta, P., Roubaud, M.C., and Torrésani, B. (2004). Blind source separation and the analysis of microarray data. J. Comput. Biol. *11*, 1090–1109.

Choi, W., Porten, S., Kim, S., Willis, D., Plimack, E.R., Hoffman-Censits, J., Roth, B., Cheng, T., Tran, M., Lee, I.L., et al. (2014). Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. Cancer Cell *25*, 152–165.

Chung, C.H., Parker, J.S., Karaca, G., Wu, J., Funkhouser, W.K., Moore, D., Butterfoss, D., Xiang, D., Zanation, A., Yin, X., et al. (2004). Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell *5*, 489–500.

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., et al. (2007). Integration of biological networks and gene expression data using Cytoscape. Nat. Protoc. *2*, 2366–2382.

Collaborative Group on Hormonal Factors in Breast Cancer (1997). Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Lancet *350*, 1047–1059.

Damrauer, J.S., Hoadley, K.A., Chism, D.D., Fan, C., Tiganelli, C.J., Wobker, S.E., Yeh, J.J., Milowsky, M.I., Iyer, G., Parker, J.S., and Kim, W.Y. (2014). Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. Proc. Natl. Acad. Sci. USA *111*, 3110–3115.

Di Palma, S., Simpson, R.H.W., Marchiò, C., Skálová, A., Ungari, M., Sandison, A., Whitaker, S., Parry, S., and Reis-Filho, J.S. (2012). Salivary duct carcinomas can be classified into luminal androgen receptor-positive, HER2 and basal-like phenotypes. Histopathology *61*, 629–643.

Dyrskjøt, L., Kruhøffer, M., Thykjaer, T., Marcussen, N., Jensen, J.L., Møller, K., and Ørntoft, T.F. (2004). Gene expression in the urinary bladder: a common carcinoma in situ gene expression signature exists disregarding histopathological classification. Cancer Res. *64*, 4040–4048.

Falcon, S., and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. Bioinformatics *23*, 257–258.

Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., and Parkin, D.M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int. J. Cancer *127*, 2893–2917.

Frigyesi, A., Veerla, S., Lindgren, D., and Höglund, M. (2006). Independent component analysis reveals new and biologically significant structures in micro array data. BMC Bioinformatics *7*, 290.

Himberg, J., Hyvärinen, A., and Esposito, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. Neuroimage *22*, 1214–1222.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. *10*, 626–634.

Jutten, C., and Hérault, J. (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. Signal Process. *24*, 1–10.

Kim, W.J., Kim, E.J., Kim, S.K., Kim, Y.J., Ha, Y.S., Jeong, P., Kim, M.J., Yun, S.J., Lee, K.M., Moon, S.K., et al. (2010). Predictive value of progression-related gene classifier in primary non-muscle invasive bladder cancer. Mol. Cancer *9*, 3.

Kong, S.L., Li, G., Loh, S.L., Sung, W.K., and Liu, E.T. (2011). Cellular reprogramming by the conjoint action of ERα, FOXA1, and GATA3 to a ligand-inducible growth state. Mol. Syst. Biol. *7*, 526.

Lee, S.-I., and Batzoglou, S. (2003). Application of independent component analysis to microarrays. Genome Biol. *4*, R76.

Liebermeister, W. (2002). Linear modes of gene expression determined by independent component analysis. Bioinformatics *18*, 51–60.

Lin, D.W., Coleman, I.M., Hawley, S., Huang, C.Y., Dumpit, R., Gifford, D., Kezele, P., Hung, H., Knudsen, B.S., Kristal, A.R., and Nelson, P.S. (2006). Influence of surgical manipulation on prostate gene expression: implications for molecular correlates of treatment effects and disease prognosis. J. Clin. Oncol. *24*, 3763–3770.

Rebouissou, S., Bernard-Pierrot, I., de Reyniès, A., Lepage, M.L., Krucker, C., Chapeaublanc, E., Hérault, A., Kamoun, A., Caillault, A., Letouzé, E., et al. (2014). EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. Sci Transl Med *6*, 244ra91.

Rouveirol, C., Stransky, N., Hupé, P., Rosa, P.L., Viara, E., Barillot, E., and Radvanyi, F. (2006). Computation of recurrent minimal genomic alterations from array-CGH data. Bioinformatics *22*, 849–856.

Sjödahl, G., Lövgren, K., Lauss, M., Patschan, O., Gudjonsson, S., Chebil, G., Aine, M., Eriksson, P., Månsson, W., Lindgren, D., et al. (2013). Toward a molecular pathologic classification of urothelial carcinoma. Am. J. Pathol. *183*, 681–691.

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA *98*, 10869–10874.

Spruck, C.H., 3rd, Ohneseit, P.F., Gonzalez-Zulueta, M., Esrig, D., Miyao, N., Tsai, Y.C., Lerner, S.P., Schmütte, C., Yang, A.S., Cote, R., et al. (1994). Two molecular pathways to transitional cell carcinoma of the bladder. Cancer Res. *54*, 784–788.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Teschendorff, A.E., Journée, M., Absil, P.A., Sepulchre, R., and Caldas, C. (2007). Elucidating the altered transcriptional programs in breast cancer using independent component analysis. PLoS Comput. Biol. *3*, e161.

Turner, R.M., Kwok, C.S., Chen-Turner, C., Maduakor, C.A., Singh, S., and Loke, Y.K. (2014). Thiazolidinediones and associated risk of bladder cancer: a systematic review and meta-analysis. Br. J. Clin. Pharmacol. *78*, 258–273.

Vallot, C., Stransky, N., Bernard-Pierrot, I., Hérault, A., Zucman-Rossi, J., Chapeaublanc, E., Vordos, D., Laplanche, A., Benhamou, S., Lebret, T., et al. (2011). A novel epigenetic phenotype associated with the most aggressive pathway of bladder tumor progression. J. Natl. Cancer Inst. *103*, 47–60.

Varley, C.L., Stahlschmidt, J., Lee, W.C., Holder, J., Diggle, C., Selby, P.J., Trejdosiewicz, L.K., and Southgate, J. (2004). Role of PPARgamma and EGFR signalling in the urothelial terminal differentiation programme. J. Cell Sci. *117*, 2029–2036.

Wilkerson, M.D., Yin, X., Hoadley, K.A., Liu, Y., Hayward, M.C., Cabanski, C.R., Muldrew, K., Miller, C.R., Randell, S.H., Socinski, M.A., et al. (2010). Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. Clin. Cancer Res. *16*, 4864–4875.